The 4th Pacific Rim Real Estate Society Conference
Perth Australia 19-21 January 1998

Information Diffusion in the Perth Housing Market

Gregory J Costello, Curtin University of Technology, Perth, Western Australia
(costellog@cbs.curtin.edu.au)

Introduction

It is often casually asserted that the housing market is inefficient. In Australia there has been very little scholarly work undertaken to test the hypothesis. The Efficient Market Hypothesis (EMH) provides a theoretical framework for analysis of how an asset's price is influenced by new information concerning the asset. Because of abundant data there have been numerous empirical tests of the EMH in securities markets. In recent years the EMH has been rigorously tested in real estate markets. The great majority of this research has occurred in North America from the period of the mid 1980's (Gatzlaff & Tirtiroglu:1994). In theory the market for any type of asset can be tested for its degree of efficiency if appropriate data exists.

This paper is a preliminary test of the EMH applied to a specific housing sub-market in Perth, Western Australia for the period 1988-1996. A repeat-sales dataset is utilised to test the random-walk properties for housing transactions during this period. The tests provide evidence of positive serial correlation for lag periods of three months and zero serial correlation for lag periods longer than six months. Some evidence is presented that there may some persistent positive serial correlation for lag periods of 4-5 months although results are still inconclusive.

Theoretical Background

Why are tests of housing market efficiency important? The foundation of finance theory is that capital markets exist to transfer funds between lenders (savers) and borrowers (producers) efficiently. In the case of real estate markets we see lending institutions transferring funds to borrowers who may or may not be engaged in productive activity. In the case of housing markets there are consumption benefits associated with ownership of housing which can be considered as the implicit rental dividend. Capital markets exist so that investors/consumers in real estate markets can borrow required funds to enable acquisition of real estate assets. Lenders with excess funds will be willing to lend if the return on funds lent exceeds the possible returns from direct investment. In this environment both borrowers and lenders in real estate markets are better off if the capital market is efficient in facilitating fund transfers. Market efficiency implies that prices for assets are accurate signals for the allocation of capital. A market is considered to be *allocationally* efficient when prices in that market are determined in a way that equates risk adjusted marginal rates of return for all producers and savers in that market. In an allocationally efficient real estate market scarce savings should be optimally allocated to investment in real estate assets so that both lenders and borrowers benefit from the exchange of funds.

This paper examines *informational* efficiency. Fama (1970) defined an efficient capital market as one in which the price of an asset fully reflects all available information. With this rigid definition there is an emphasis on the relationship between market prices and information. In securities markets this emphasis requires that the reaction of prices to new information is both "instantaneous" and unbiased. This requirement for "instantaneous" diffusion of information implies that if one part of the market is uninformed concerning information which influences the price of an asset, then the part of the market with the information can trade on this information to earn "excess" or "abnormal" profits. If such profits are possible then the market is inefficient. A biased price reaction can be considered an overreaction or an under-reaction to the arrival of new information that might affect the price of an asset. In an efficient market, competition between informed buyers and sellers should fully reflect the available information set, therefore there can be no biased price reactions.

The EMH can be formally defined and tested using the martingale model. This is an application of the fair-game model. A stochastic process, x_t , is termed a martingale with respect to the sequence of information sets, Φ_t , if:

$$E_t(x_{t+1} \mid \Phi_t) = x_t \qquad (1)$$

where E_t denotes the expectations operator conditional on the information set Φ_t available at time t. This indicates that the best forecast of x_{t+1} is x_t, given the relevant information set Φ_t (note that x_t is assumed to be in Φ_t). The martingale model of efficiency implies that the market is in equilibrium and that investors cannot consistently earn above-normal returns on investments based on any information set, Φ_t.

The EMH framework is closely aligned to the perfect competition model from microeconomics. Under perfect competition it is assumed that there are sufficient buyers and sellers in a market so that each participant cannot individually determine market prices. Prices reflect consensus opinions about the market value of assets although participants may not have homogenous expectations about the future benefits of ownership of the assets. Perfect competition and market efficiency assume free entry and exit of resources and participants with no capital or information barriers to entry to the market. Market participants should have equal access to all information likely to affect the future benefits associated with ownership of individual assets. Capital markets do not completely satisfy all of the assumptions of perfect competition. Most capital markets have some imperfections that inhibit the flow of capital and/or information and reduce market efficiency below the perfect competition optimum. In this sense market efficiency represents a relative rather than an absolute concept.

In real estate markets these market imperfections are characterised by barriers to entry and informational inefficiencies. The local nature of real estate markets creates high information (search) costs. High information costs do not necessarily create an inefficient market. These costs create the requirement for higher returns to the investment to compensate for the higher information costs. Real estate transactions are also characterised by high levels of transaction costs. In a similar manner to high information costs transaction costs could reduce arbitrage movements of real investors between specific property types and geographic markets. Capital constraints are also significant features of real estate markets. Information dissemination in markets may be constrained by limited investor participation in certain market segments. Under these conditions it is likely that market segmentation exists with differing levels of efficiency in segments of real estate markets. This is an empirical issue.

Linneman (1986) suggests that housing markets are characterised by highly dispersed information sets. An implicit assumption of the EMH is that all information is available to all investors at the same cost. The reality of housing markets is that the participants may be experienced or inexperienced in housing transactions. They may be investors or consumers (owner-occupiers) of housing. They have many varied housing budgets and location preferences resulting in many different information sets. Since real estate prices are usually the product of bilateral negotiations between buyers and sellers information is extremely valuable. Wheaton (1990) suggests that over time households experience stochastic demographic changes which mismatch some households with their current house. This process generates a desired sale and new purchase for a mismatched household. These households undertake a costly search among houses for sale, deciding whether to negotiate for a given unit or continue the search process. These market participants learn about market fundamentals from the observation of housing units for sale, information from brokers and by negotiation with sellers. Sellers will learn in the same way and by observing the behavior of potential buyers. Transaction prices of comparable units provide important information used by both sides of the market to set parameters for negotiation. Since the information set used in any negotiation is necessarily lagged, it must be inferred that lagged price movements will influence current prices in an efficient housing market.

The question of the appropriate lag period that is consistent with real estate market efficiency is important and is also an empirical issue. Unlike security markets where stock prices are available daily, real estate prices are subject to considerable delay before becoming available as public information. Real estate transactions are negotiated with a settlement (closing) period occurring between date of transaction and final settlement. In Western Australia full public information is only available after settlement, typically involving a lag period of 60-90 days from negotiation date of the transaction. In this environment of information diffusion it is possible to hypothesise that there are in fact several relevant information sets and time lags appropriate to tests of the EMH in housing markets. One information set is the "local" information set consisting of shorter lag periods. This is a "word of mouth" process of information diffusion that is very important to active buyers and sellers. Another information set is the "full" information set consisting of a longer lag period. This corresponds with settlement of a transaction and full public details of the transaction (including hedonic characteristics) being available. This is the information set desirable for portfolio investors looking to exploit any systematic inefficiency in a housing market for abnormal returns.

Gau (1987) suggests that a real estate market is efficient if despite market imperfections, value-influencing information is effectively capitalised into market prices. In such a market, participants act on information received and prices adjust quickly upon the arrival of information. The adjustment of prices occurs at a rate so that individual investors trading on the information are unable to earn abnormal returns. Real estate market efficiency should not be viewed as an absolute concept, rather more one of degree. Given real estate market imperfections, it is likely that real estate markets are not as efficient as securities markets and that some regional real estate markets are more or less efficient than others. It is also feasible that real estate markets capitalise different types of information at different rates and in differing degrees. Some information may be fully capitalised whereas prices may fail to reflect some other sets of information.

Data

The data used in this study has been obtained from the Western Australian Valuer General's Office (VGO). From this data hedonic and repeat-sale indexes were constructed. Although only repeat-sales methods are used in the majority of this study some hedonic results are reported for comparison purposes. The Australian Bureau of Statistics (ABS) provided consumer price index (CPI) numbers. The final repeat-sales index was constructed from 18,583 observations.

The initial VGO data file comprised the population of all transfers of residential strata title property in the Perth metropolitan area for the period of September 1988 – January 1996. This file recorded transaction details for 49,558 *individual* properties. It is important to understand the three potential types of observations contained in the raw data from which the dataset was constructed. The observation types are shown as Exhibit 1.

Exhibit 1: Transaction Types

	SALE1	DATE1	SALE2	DATE2	SALE3	DATE3
Type 1	56000	19960124	0	0	0	0
Type 2	105500	19960730	104000	19930529	0	0
Type 3	133500	19960802	124000	19931006	118000	19910111

- *Type 1*: Those observations which sold only once during the sample period 1988-96. None of these observations appear in the repeat-sales sample
- *Type 2*: Those observations which sold only twice during the sample period 1988-96. These observations appear in the repeat-sales sample as a single repeat-sale.
- *Type 3*: Those observations which sold three *or more* times during the sample period 1988-95.[1] These observations appear in the repeat-sales sample as two repeat-sales.

Exhibit 2: Descriptive Statistics for Transaction Types

Observation Type	Number of Observations	Median Sale Price - Most Recent Sale	Median Building Area SQM	Median Building Age at - Most Recent Sale
Type 1	32,173	97,500	90	10
Type 2	13,354	91,275	82	14
Type 3	4,031	85,000	74	18

Some descriptive statistics for each of these observation types are shown as Exhibit 2. There are some significant relationships between frequency of sale and important hedonic characteristics. The observation types which sell most frequently (Type 3) are

[1] The VGO data provides only the three most recent sales. This censors the sample from potential sale_4...sale_n type observations.

older, smaller, and cheaper. Furthermore, this trend appears to be hierarchical in that Type 1 observations are the most expensive sales, the largest buildings and the youngest in age of building. These statistics are supportive of the *starter home hypothesis* whereby first home owners buy the cheapest properties and sell (trade-up) more frequently. This indicates that the repeat-sales index is a biased sample of Perth housing units. This is intentional, because data in this study has been selected in an attempt to maximise index temporal accuracy. This is discussed further under Index Accuracy below.

Appendix 1 summarises hedonic characteristics for *all transactions* (type 1,2,3) and individual quarterly periods. It is evident that there has also been some significant temporal variation in important hedonic characteristics. Median selling prices have increased from $65,000 in the base index period to $109,250 in the final sample period. This represents an increase of 68%. During the same period median building area increased approximately 15%. Building ages at sale became slightly higher during the sample period. This evidence of varying quality of housing units through the sample period is indicative of problems associated with index methods using simple measures of central tendency such as median sale price indexes.

Index Methodology

The index type used for this paper is the Weighted Repeat Sales (WRS) method as used by Case & Shiller (1989). Hedonic index techniques were also used but are not reported as the results are consistent with the WRS results. Exhibit 1 indicates that type 2 and type 3 variables in original VGO format can be transformed to yield a repeat-sales dataset.

The methodology for construction of a repeat-sales index was pioneered by Bailey Muth and Nourse (1963) (BMN). The essential data required for a single property to be used in a repeat-sales index is an initial sale and date and a subsequent sale and date. Due to the same property transacting in different time periods it is assumed that property attributes remain unchanged and the resultant price difference is due to the intervening time period. The repeat-sales technique avoids many of the problems associated with hedonic explicit-time models. It can be shown that the repeat-sales model is a variant of the explicit-time variable approach by contrasting equation 2 below with equation 3.

$$\ln P_{it} - \ln P_{i\tau} = \left(\sum_{j=1}^{k} \beta_j \ln X_{jit} + \sum_{t=1}^{T} c_t D_{it} \right) - \left(\sum_{j=1}^{k} \beta_j \ln X_{ji\tau} + \sum_{\tau=1}^{T} c_\tau D_{i\tau} \right) + \varepsilon_{it\tau} \qquad (2)$$

where P_{it} and $P_{i\tau}$ are the prices of repeat-sales, with the initial sale at time τ and the second sale at time t; X_τ and X_t denote the structural and locational attributes at each respective sale, c_τ and c_t ,the time coefficients of D_{it} and $D_{i\tau}$. If it is assumed that the quality of the housing unit is constant between transactions then the difference between transaction prices at the two dates can be considered as a function solely of the time period, equation (2) reduces to:

$$\ln P_{it} - \ln P_{i\tau} = \sum_{t=1}^{T} c_t D_{it} - \sum_{\tau=1}^{T} c_\tau D_{i\tau} + \varepsilon_{it\tau} \qquad (3)$$

The dependent variable becomes the logarithm of the price ratio from the property having sold twice. The log price relatives are then regressed on a set of dummy variables corresponding with the time periods. The estimating equation becomes:

$$\ln\left(\frac{P_{it}}{P_{i\tau}}\right)=\sum c_t D_{it}+\varepsilon_{j\tau t} \tag{4}$$

where $P_{it} / P_{i\tau}$ is the price relative for property *i*; D_{it} is a dummy variable which equals -1 at the time of initial sale and +1 at the time of the second sale, and 0 otherwise; c_t is the logarithm of the cumulative price index in period *t*; and $\varepsilon_{j\tau t}$ is a disturbance term. The logarithm of the initial value of the index is normalised by setting initial values in D_1 equal to zero, (i.e. omitting base period time category) and the *T* subsequent coefficients are estimated by OLS regression (Gatzlaff & Ling:225).

The most discussed problem of the Repeat-sales method is the issue of sample selection bias. This has been discussed in the context of the "starter home hypothesis" where houses that sell frequently may be starter homes bought by individuals with a short expected duration of stay. This type of house will therefore appear relatively frequently in sales of sold houses. Another significant sample selection bias problem is heteroscedasticity due to the influence of omitted variables such as holding period between transactions. It can been shown that holding periods are not uniformly distributed through the sample period. Specifically, short holding periods are under represented in the beginning and end periods of the index. This problem of holding period induced heteroscedasticity in BMN regressions has been addressed by several authors. Case & Shiller (1989) used a three step Weighted (Generalised) Least Squares correction to estimate a BMN index with observations weighted according to holding period (Weighted Repeat-Sales). They model the WRS method on the assumption that the log price P_{it} of the *i*th house at time t is given by:

$$P_{it}=C_t+H_{it}+N_{it} \tag{5}$$

where P_{it} is the log price of the *i*th house at time *t*, C_t is the log of the city wide level of housing prices at time *t*, H_{it} is a Gaussian random walk (where ΔH_{it} has zero mean and variance σ_h^2) that is uncorrelated with C_t and H_{jt} $i \neq j$ for all *T*, and N_{it} is an identically distributed normal noise term (which has zero mean and variance σ_N^2) and is uncorrelated with C_t and H_{jT} for all *j* and *T* with N_{jT} unless $i = j$ and $t = T$. Here, H_{it} represents the drift in individual housing value through time, and N_{it} represents the noise in price due to imperfection in the market for housing. The introduction of the two noise terms is in recognition of heterogeneous characteristics of real estate markets and market imperfections in the selling process such as the random arrival of interested purchasers. In these circumstances the final sale price may not be identical to true value.[2] The WRS index utilises a three-step weighted (generalised) least squares procedure. In the first step the BMN procedure is followed exactly, and a vector of regression residuals is calculated. In the second step the squared residuals from the first

[2] This recognition of two noise terms has some important "errors in variables" effects which will influence tests of market efficiency using autocorrelation techniques on lagged differences in index values.

step are regressed on a constant and the time interval between sales (holding period). The constant term is the estimate of σ_N^2, the noise in price due to imperfections in the market for housing. The slope term is the estimate of σ_H^2, the drift in individual housing unit value through time. In the third step a generalised least squares regression (weighted regression) is run by first dividing each observation in the step-one regression by the square root of the fitted value in the second-stage regression and running the regression again. Index numbers for quarterly periods are attached as Appendix 2. And represented graphically as Chart 1.

Index Accuracy

The data for this study was selected with the intention of maximising the accuracy of measurement of index period differences for the purpose of conducting serial correlation tests using lagged index differences. Case & Shiller (1989) proposed several methods of describing how well the index variables are measured. One method involves computing the ratio of the standard deviation of a variable to the average standard error for that variable. Exhibit 3 summarises this approach with a variety of index methods using OLS procedures for Perth data.

Exhibit 3 Index Accuracy

Index Type	Ratio - Log index levels	Ratio - Log index first difference
Simple Hedonic Quarterly Periods	10.23	4.13
Complex Hedonic Quarterly Periods	14.92	5.43
Simple Hedonic Monthly Periods	5.71	1.28
Complex Hedonic Monthly Periods	8.25	1.60
Weighted Repeat Sales Quarterly Periods	16.03	5.84
Weighted Repeat Sales Monthly Periods	8.70	1.55

Higher ratios indicate more accurately measured index characteristics. It can be seen above that for the log index in levels quarterly levels are more accurately measured than monthly differences and the weighted repeat-sales (WRS) method yields the highest ratio. For first differences the quarterly periods are most accurately measured with the WRS method. With monthly first differences the complex hedonic model yields the most accurate measurement of monthly differences although the difference between the WRS ratio is minimal. Case & Shiller (1989) describe similar figures on the log index in levels as "accurate", and ratios in the vicinity of 2.7 – 4.0 for *annual* differences as "fairly accurate". Ratios in the vicinity of 1.0 – 2.0 for quarterly differences were discussed: "we thus cannot accurately describe the quarterly changes in the log prices, though the index will give a rough indication." Case & Shiller (1989:127). This indicates that for Perth data the quarterly first differences for all index methods are measured quite accurately with the ratios higher than for some annual differences in Case & Shiller's (1989) US study. The monthly indexes are useful in providing a rough indication and providing more observations on second and third differences than would be provided by a quarterly index. It will be shown that when second differences of the monthly WRS index are taken the accuracy improves significantly.

Another useful diagnostic for assessing index accuracy is shown in Exhibits 6 and 7. This is the regression for an individual difference (1st, 2nd, 3rd) of real log index A on the contemporaneous difference change in index B (with the corresponding reverse regressions not reported). The coefficient should be 1.00 if the indexes are measured perfectly but will vary from one due to the errors-in-variables problem. Exhibit 6 shows the estimated coefficient for quarterly index first differences is .955 with R^2 = .91. This is an indication that quarterly first differences are well measured. The estimated coefficient for monthly index first differences (Exhibit 7) is .483 with R^2 = .183. This indicates monthly first differences are not well measured. The estimated coefficient for monthly index second differences (Exhibit 7) is .792 with R^2 = .501. This indicates monthly second differences are more accurately measured than first differences. The estimated coefficient for monthly index third differences (Exhibit 7) is .916 with R^2 = .730. This corresponds with the quarterly index diagnostic (Exhibit 6) confirming that quarterly or monthly index third differences are accurately measured.

Seasonality

Exhibit 4 below gives sample statistics for quarterly differences of $W(t) = WRS(t) - \ln(CPI(t))$. This is the WRS log index deflated by the contemporaneous CPI index. The growth in real prices for the index period was approximately one third of one per cent per quarter or 1.12% per annum. The standard deviation in quarterly real price changes is slightly higher than three percent per quarter.

Exhibit 4: Seasonality Statistics

Quarterly Changes in Real WRS Log Price Index: z = W(t) - W(t-1)					
	Mean z for Quarter t				H_o: All Quarters Same Mean
All Quarters Mean z std.z	t = 1 (t-stat)	t = 2 (t-stat)	t = 3 (t-stat)	t = 4 (t-stat)	F Prob.
0.0028 0.0312	0.0154 0.9779	-0.0061 -0.7047	-0.0102 -1.0230	0.0095 0.5238	1.1678 0.3410

First quarter changes appear to be high and third quarter changes low although seasonality is not statistically significant. Individual quarterly differences were subjected to a pooled-variance *t* test with the null hypothesis being that the mean difference for any quarter was not different from the mean difference for all quarters. The null hypothesis could not be rejected for any quarterly period. A one way analysis of variance was used including all quarterly periods. The null hypothesis was that all quarters had the same mean. It was not rejected at the .05 level of significance.

The Distribution of Index Changes

The variable studied in Exhibit 5 (excepting !) denoted x_t is the first difference of the respective index type. For the log WRS index, $x_t = WRS_t - WRS_{t-1}$. This variable is used because the change in log prices is the yield under continuous compounding, $P_{t+1} = P_t \exp(\ln P_{t+1} - \ln P_t)$. Also the variability of non-logarithmic price changes in an index is an increasing function of the scale of the index numbers. The use of logarithms neutralises this influence.

The WRS variables are first differences for the WRS index. The *W(t)* variables are first differences for the real log price index. This is the WRS index deflated by the contemporaneous CPI index, $W(t) = WRS(t) - \ln(CPI(t))$. The official CPI numbers are only available for quarterly periods necessitating monthly indexes be deflated using regular intervals for monthly periods between the known quarterly CPI index numbers.

Exhibit 5: Index Differences Descriptive Statistics

Parametric Statistics								Non-Parametric Tests	
Index Type	N	Mean	Median	Mode	Std. Deviation	Skewness	Kurtosis	Kolmogorov-Smirnov Z	Runs Z
WRS Quarterly	30	0.0117	0.0077	-0.0319	0.0300	1.8048	4.7430	1.1582	-2.4155*
W(t) Quarterly	30	0.0028	0.0011	-0.0521	0.0312	1.0276	2.5435	1.0007	-2.4155*
WRS Monthly	91	0.0044	0.0053	-0.0355	0.0154	0.5353	1.4546	1.0726	1.7934
W(t) Monthly	91	0.0014	0.0028	-0.0425	0.0156	0.1566	1.1096	0.8586	0.5283
!2 *W(t)* Monthly	90	0.0023	0.0026	-0.0625	0.0243	0.6281	1.8487	1.0905	-3.18***
!3 *W(t)* Monthly	89	0.0029	0.0046	-0.0779	0.0331	0.8739	2.5485	1.33**	-5.22***

!2 second difference
!3 third difference
* statistically significant at .02 level.
** statistically significant at .06 level.
*** statistically significant at any level.

These statistics test the distributions of index differences against the null hypothesis of the distributions constituting a random walk. The random walk hypothesis preceded the EMH. In testing for market efficiency there is an assumption that if a market is efficient the differences in a log index constitute a random walk. The sequence of random variables $(x_t, t = 1,2,...)$ is called a random walk if the increments $x_t - x_{t-1} = e_t$ are independently distributed. The random-walk model assumes that successive percentage changes in an asset's price are independent and are identically distributed over time. A random walk model does not by itself imply market efficiency.

The parametric statistics above indicate parameters associated with a normal distribution. A uniform normal distribution is characterised by skewness = 0 and kurtosis = 3. All distributions above are positively skewed and have higher peaks than the parameters associated with a normal distribution. Charts 2-4 attached provide histograms for some of these distributions. The histograms of W(t) variables for second and third differences on the monthly index series provide more observations to assess the distribution. Descriptively the distributions appear to be quite symmetrical with some positive skewness. There is more probability than normal within ± ½ standard deviation, and less probability between ± 1-2 standard deviations. There is more probability at greater than +2 ½ standard deviations. In brief the distributions have a fat positive tail, high peaked centres, and are hollow in between. This indicates that positive changes in the index tend to be infrequent but of greater magnitude than negative changes. This is supported by all index differences having positive mean/median differences and negative modes.

The non parametric statistics above provide alternative tests for normality. The Kolmogorov-Smirnov test compares the observed cumulative distribution function for index differences against the null hypothesis that the sampling distribution of differences is a normal distribution. The null hypothesis of a normal distribution cannot be rejected for any of the distributions above. The Runs test analyses the sign of index changes during the sample period to test for independence of the changes in prices for the individual periods. The term *runs* refers to consecutive periods of price changes of the same sign. If it is assumed that price changes in individual periods are independent then the expected number and length of runs can be calculated. By comparing the actual number of runs with the expected numbers of runs evidence of dependence in price changes between periods can be gathered. The above results reject independence for the quarterly index differences. Independence cannot be rejected for the monthly series but it must be remembered that these differences are not accurately measured so apparent independence is probably due to the greater noise in the monthly index. When more accurate second and third differences are taken for the monthly series the null hypothesis of independence of index period differences is rejected with statistical significance at any level.

In summary these results provide evidence to reject a random walk hypothesis. It appears that there is significant dependence between index periods. Further details as to the characteristics of this dependence are indicated by results from the following serial correlation tests.

Serial Correlation Tests

A standard method for testing any random walk property of prices is to regress the change in the index on lagged changes in the index and to test for serial correlation. This ignores the problem of noise in any estimated index that can cause spurious correlation. The presence of errors in variables creates noise in the estimated WRS index. Case & Shiller (1989) demonstrate this with several examples attached as Appendix 3. They demonstrate that it is inappropriate to test efficiency of a housing market by regressing real changes in the WRS index onto lagged changes and testing for significance of the coefficients. The same noise in individual house sales contaminates both dependent and independent variables. A simple expedient for this problem is to split the sample of individual house sales into two individual random samples and then estimate new WRS indexes.

Another problem is created if overlapping data is used in OLS regression procedures. Overlapping data occurs where the index difference used is greater than the lag period. For example if a monthly index is used and the third difference is taken and lagged by one month. In this case third differences are systematically overlapping each other by one month at each lag. This causes a violation of the OLS assumption of independence of the error terms. With overlapping data successive error terms are not independent causing problems of efficiency with the estimators and hence problems with tests of statistical significance using t and F test distributions.

This is a problem because test methodology can be greatly improved by using overlapping data. It provides more observations for testing varying difference periods with single period lags. Case & Shiller (1989) suggest a correction procedure as used by Hansen & Hodrick (1980) and Clapp Dolde & Tirtiroglu (1995) demonstrate a variation of the method. The correction procedure has not been applied in the following tests. For this reason non-overlapping data is used. Where benefit can be gained from considering an overlapping period, R^2, F and t statistics are not recorded.

Following Case & Shiller's (1989) procedure the repeat-sales data was split evenly with a random allocation of houses into either sample A or B. Two log price indexes WRS_A and WRS_B were estimated using the new samples. The random walk properties of the index can then be tested by regressing changes in the index WRS_A on lagged changes in the real index WRS_B. Both sides of the equation are contaminated by noise, but since the same house does not appear in the index on both sides of the equation the noise terms are not correlated. With this methodology, the null hypothesis that $\beta_1 = 0$ for specified lag periods can be tested. If the β_1 coefficients are not equal to zero and statistically significant there may be evidence to reject a random walk hypothesis of weak-form market efficiency. The relevant lag period is very important. As hypothesised earlier because participants in housing markets must use lagged information sets of past prices it is expected that in a weak-form efficient housing market positive serial correlation will be observed for shorter lag periods i.e. $\beta_1 > 0$

These regressions are shown below in Exhibits 6 & 7. The first regressions for changes of real log index A on the contemporaneous quarterly change in index B where $L = 0$ has been completed as a diagnostic on the methodology. It is also a useful indication as to how well specific index differences are measured. The β_1 coefficient should be 1.00 if the indexes are measured perfectly but will vary from one due to the errors-in-variables problem. The estimated coefficients for quarterly first differences and monthly third differences are very close to 1.00.

The results displayed in Exhibit 6 confirm significant positive serial correlation for a lag period of one month (regressions 2&3). Lag periods from 2-4 quarters (regressions 4,5,6) indicate no significant serial correlation. More information is available with the monthly index series provided in Exhibit 7.

Using monthly first differences it is evident that significant positive serial correlation is present for lag periods of up to three months (regressions 2-5). Both regressions for the lag period of three months are reported. Regression 4 (Sig .078) indicates that the null hypothesis of zero serial correlation could not be rejected at .05 significance level. The corresponding reverse regression 5 supports rejection of zero serial correlation. For all other lag periods the results of the first regression are supported by results for the corresponding reverse regression therefore the reverse regression results are not reported. Regression 6 is the regression for monthly first differences lagged four months. The coefficient is positive but not statistically significant. Additional regressions (not reported) for up to twelve month lag periods were taken with none of the results confirming statistically significant positive or negative serial correlation.

Exhibit 6: Regression of Changes in Quarterly Real Log Index Estimated with One Half of Sample on Changes in Real Log Index Estimated with Other Half of Sample.

$$W_j(t) - W_j(t - \Delta_t) = \beta_0 + \beta_1(W_k(t - L) - W_k(t - \Delta_t - L)) + u(t)$$

Regression	Parameters: Δ_t = Index period difference L = Lag period	No. obs. S.E.E.	β_0 *(t)* Sig	β_1 *(t)* Sig	R^2 Adj R^2
1	$J = A, k = B, \Delta_t = 1, L = 0$	29 0.010	0.001 *(0.377)* 0.709	0.955 *(16.848)* 0.000	0.910 0.907
2	$J = A, k = B, \Delta_t = 1, L = 1$	28 0.024	-0.001 *(-0.251)* 0.803	0.478 *(3.430)* 0.002	0.304 0.278
3	$J = B,1\ k = A, \Delta_t = 1, L = 1$	28 0.022	-0.002 *(-0.422)* 0.676	0.561 *(4.270)* 0.000	0.403 0.381
4	$J = A, k = B, \Delta_t = 1, L = 2$	27 0.022	-0.003 *(-0.734)* 0.469	0.023 *(0.172)* 0.865	0.001 -0.037
5	$J = A, k = B, \Delta_t = 1, L = 3$	26 0.023	-0.003 *(-0.639)* 0.528	-0.017 *(-0.126)* 0.900	0.001 -0.039
6	$J = A, k = B, \Delta_t = 1, L = 4$	25 0.021	-0.001 *(-0.236)* 0.815	-0.005 *(-0.039)* 0.969	0.000 -0.042

Regressions 7-11 use monthly second differences. It is evident by comparing regression 7 with regression 1 that the second difference is more accurately measured. Regressions 8 & 9 confirm significant positive serial correlation is present for lag periods of up to four months. The coefficients also indicate greater serial correlation for the shorter lag period of two months than for four months. Regression 11 for a lag period of six months indicates that there is no statistically significant positive serial correlation present. The overlapping regression 10 for a lag period of 5 months confirms a positive coefficient but tests of statistical significance are invalidated (discussed above) and are therefore not reported.

Regressions 12-17 use monthly third differences. It is evident in comparing regression 12 with regressions 7 & 1 that the third difference is more accurately measured than first or second differences. Regressions 14 & 17 confirm previous results of significant positive serial correlation for a lag period of three months with no statistically significant positive serial correlation for a lag period of six months. Regressions 13, 15, 16 are for overlapping lag periods of 1, 4 and 5 months respectively. The coefficients confirm positive serial correlation but tests of statistical significance are not reported. Further regressions (not reported) for lag periods of up to twelve months were completed and confirm previous results of no statistically significant positive or negative serial correlation.

Exhibit 7: Regression of Changes in Monthly Real Log Index Estimated with One Half of Sample on Changes in Real Log Index Estimated with Other Half of Sample.

$$W_j(t)-W_j(t-\Delta_t)=\beta_0+\beta_1(W_k(t-L)-W_k(t-\Delta_t-L))+u(t)$$

Regression	Parameters: Δ_t = Index period difference L = Lag period ! = Overlapping data	No. obs. S.E.E.	β_0 *(t)* Sig	β_1 *(t)* Sig	R^2 Adj R^2
1	$J = A,\ k = B,\ \Delta_t = 1,\ L = 0$	90 0.018	0.001 *(0.331)* 0.741	0.483 *(4.463)* 0.000	0.183 0.174
2	$J = A,\ k = B,\ \Delta_t = 1,\ L = 1$	89 0.018	0.000 *(0.229)* 0.820	0.373 *(3.341)* 0.001	0.113 0.102
3	$J = B,\ k = A,\ \Delta_t = 1,\ L = 2$	88 0.016	0.000 *(0.289)* 0.773	0.217 *(2.453)* 0.016	0.065 0.054
4	$J = A,\ k = B,\ \Delta_t = 1,\ L = 3$	87 0.019	0.000 *(0.032)* 0.974	0.207 *(1.781)* 0.078	0.036 0.024
5	$J = B,\ k = A,\ \Delta_t = 1,\ L = 3$	87 0.016	0.000 *(-0.011)* 0.992	0.228 *(2.676)* 0.009	0.077 0.066
6	$J = B,\ k = A,\ \Delta_t = 1,\ L = 4$	86 0.016	0.000 *(-0.107)* 0.915	0.089 *(1.013)* 0.314	0.012 0.000
7	$J = A,\ k = B,\ \Delta_t = 2,\ L = 0$	89 0.020	0.001 0.320 0.750	0.792 9.392 0.000	0.501 0.495
8	$J = A,\ k = B,\ \Delta_t = 2,\ L = 2$	87 0.023	0.000 -0.052 0.959	0.573 5.858 0.000	0.285 0.277
9	$J = B,\ k = A,\ \Delta_t = 2,\ L = 4$	85 0.021	-0.001 -0.483 0.630	0.216 2.633 0.010	0.076 0.065
10	! $J = A,\ k = B,\ \Delta_t = 2,\ L = 5$	84 0.023	-0.002	0.203	
11	$J = B,\ k = A,\ \Delta_t = 2,\ L = 6$	83 0.019	-0.002 -1.194 0.236	0.057 0.778 0.439	0.007 -0.005
12	$J = A,\ k = B,\ \Delta_t = 3,\ L = 0$	88 0.019	0.000 0.115 0.909	0.916 15.344 0.000	0.730 0.727
13	! $J = A,\ k = B,\ \Delta_t = 3,\ L = 1$	87 0.020	0.000	0.836	
14	$J = B,\ k = A,\ \Delta_t = 3,\ L = 3$	85 0.025	-0.002 -0.612 0.542	0.420 5.681 0.000	0.278 0.269
15	! $J = A,\ k = B,\ \Delta_t = 3,\ L = 4$	84 0.027	-0.003	0.356	
16	! $J = B,\ k = A,\ \Delta_t = 3,\ L = 5$	83 0.024	-0.003	0.138	
17	$J = A,\ k = B,\ \Delta_t = 3,\ L = 6$	82 0.026	-0.004 -1.350 0.181	0.113 1.325 0.189	0.021 0.009

Conclusions

These results confirm the existence of positive serial correlation of information diffusion in Perth housing markets for periods of up to three months. It is also evident that stronger positive serial correlation applies to shorter lag periods. This may provide evidence for rejection of a weak-form efficient housing market although at this time the results are inconclusive. The issue of an appropriate lag period where information sets can be considered as “available” is important. In Western Australia the lag period of three months corresponds approximately with the lag period from date of transaction until full transaction information is published. As discussed previously this constitutes the “full” information set that would be used by portfolio investors seeking to exploit systematic inefficiencies in housing sub-markets. If lag periods of longer than three months display no significant serial correlation then it is likely that these results are consistent with a weak-form efficient housing market. The evidence presented here with overlapping regressions indicates that some positive serial correlation may exist for lag periods of four and five months. Further work with an alternative methodology is required to validate the statistical significance of these results. In addition, segmentation of the sample to examine whether different pricing levels, property types or locations are influential in explaining patterns of serial correlation would be instructive. Finally, full rejection of the EMH requires that observed inefficiencies in a market can be exploited to achieve “abnormal” returns. It is necessary to investigate this possibility further with the inclusion of appropriate transaction costs before rejecting the hypothesis of a weak-form efficient Perth housing market.

References

Bailey, M. J., Muth, R. F., Nourse, H. O. (1963). "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association*. 58: 933-942.

Case, K. E., Shiller, R.J. (1989). "The Efficiency of Market for Single-Family Homes," *The American Economic Review* 79, 125-137.

Clapp, J.M., Dolde, W., Tirtiroglu, D. (1995). "Imperfect Information and Investor Inferences from Housing Price Dynamics," *Real Estate Economics*, Fall 1995, Vol 23 No 3, 239-269.

Fama, E.E (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* 25, 383-420.

Gatzlaff, D. H., Ling, D.C. (1994). "Measuring Changes in Local House Prices: An Empirical Investigation of Alternative Methodologies," *Journal of Urban Economics.* 35: 221-244.

Gatzlaff, D.H., Tirtiroglu, D. (1995). "Real Estate Market Efficiency: Issues and Evidence" *Journal of Real Estate Literature*, Vol 3, No 2, pp 157 189.

Gau, G.W. (1987). "Efficient Real Estate Markets: Paradox or Paradigm?" *Journal of the American Real Estate and Urban Economics Association* 15, 1-15.

Hansen, L.P., Hodrick, J.R. (1980). "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, October 1980, 88, 829-53.

Linneman, R (1986). "An Empirical Test of the Efficiency of the Housing Market," *Journal of Urban Economics* 20, 140-154.

Wheaton, W. C., (1990). "Vacancy, Search, and Prices in a Housing Market Matching Model," *Journal of Political Economy* 98(6):1270-1292.

Appendix 1: Hedonic Variables – Descriptive Statistics

		Sale Price - $'000			Building Area - SQM			Age of Building - Years		
Quarter	N	Mean	Median	Std dev	Mean	Median	Std dev	Mean	Median	Std dev
88:Q3	2210	74.1	65.0	47.5	82	78	30	13	12	10
88:Q4	2917	80.7	73.0	45.0	81	76	30	13	12	9
89:Q1	2708	90.7	82.0	49.3	81	77	30	13	13	9
89:Q2	1586	91.8	82.0	52.6	80	76	29	14	13	11
89:Q3	1662	95.7	85.0	56.2	85	80	32	13	12	11
89:Q4	1588	98.5	85.5	59.6	86	82	30	12	11	11
90:Q1	2094	95.9	85.0	54.0	85	82	30	13	11	11
90:Q2	1614	97.4	85.8	50.3	88	84	31	12	11	11
90:Q3	1961	94.3	84.0	47.0	89	85	30	12	10	11
90:Q4	1922	93.9	84.0	52.0	87	84	31	12	10	12
91:Q1	2120	92.3	81.5	47.7	87	82	30	13	11	11
91:Q2	2172	95.3	83.3	54.1	88	84	31	13	11	11
91:Q3	1833	94.5	83.0	50.9	88	85	30	13	11	12
91:Q4	1391	93.7	83.0	53.6	86	82	30	14	12	12
92:Q1	2148	94.5	85.0	48.9	88	84	30	14	13	11
92:Q2	2261	98.0	86.0	58.2	89	85	31	13	12	12
92:Q3	2240	97.5	85.0	57.8	89	84	32	14	12	12
92:Q4	2112	103.2	88.5	66.1	90	85	32	14	12	12
93:Q1	2483	105.6	92.5	69.3	91	86	32	15	13	13
93:Q2	2592	105.3	92.5	56.2	90	86	32	15	13	13
93:Q3	2993	107.9	94.5	60.0	91	86	32	14	12	12
93:Q4	2839	114.3	98.5	69.1	92	88	33	14	13	12
94:Q1	3810	116.9	102.5	68.0	91	87	34	15	14	12
94:Q2	2778	119.4	105.0	71.9	90	86	33	15	14	13
94:Q3	2659	125.9	112.0	76.7	93	88	34	14	13	12
94:Q4	2223	124.0	108.0	73.5	91	86	34	15	14	13
95:Q1	2237	122.3	108.0	69.7	91	87	33	16	15	13
95:Q2	2251	129.6	115.0	80.8	95	90	35	15	14	13
95:Q3	2367	126.9	111.0	77.7	96	90	37	14	14	12
95:Q4	1916	128.2	112.0	75.6	94	90	35	15	14	12
96:Q1	1112	119.4	109.3	67.1	90	86	31	17	16	12

Appendix 2 Index Numbers for Quarterly Periods

Quarter	LN_WRS	LN_WRS_A	LN_WRS_B	LN_CPI	W(t)	WRS	REAL_WRS	CPI
88:Q3	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000
88:Q4	0.1050	0.1075	0.1022	0.0209	0.0841	1.1107	1.0877	1.0211
89:Q1	0.2071	0.2092	0.2051	0.0307	0.1764	1.2301	1.1929	1.0311
89:Q2	0.2246	0.2235	0.2253	0.0520	0.1726	1.2518	1.1884	1.0534
89:Q3	0.1955	0.1941	0.1966	0.0750	0.1205	1.2159	1.1281	1.0779
89:Q4	0.2114	0.2095	0.2136	0.0954	0.1159	1.2354	1.1229	1.1001
90:Q1	0.2129	0.2130	0.2129	0.1184	0.0945	1.2372	1.0991	1.1257
90:Q2	0.1812	0.1784	0.1836	0.1351	0.0461	1.1986	1.0472	1.1446
90:Q3	0.1493	0.1516	0.1467	0.1428	0.0065	1.1610	1.0065	1.1535
90:Q4	0.1403	0.1484	0.1308	0.1666	-0.0264	1.1506	0.9740	1.1813
91:Q1	0.1445	0.1437	0.1456	0.1572	-0.0126	1.1555	0.9875	1.1702
91:Q2	0.1419	0.1370	0.1468	0.1562	-0.0143	1.1525	0.9858	1.1691
91:Q3	0.1449	0.1442	0.1456	0.1619	-0.0170	1.1559	0.9831	1.1758
91:Q4	0.1550	0.1531	0.1570	0.1657	-0.0107	1.1676	0.9894	1.1802
92:Q1	0.1520	0.1524	0.1517	0.1657	-0.0137	1.1642	0.9864	1.1802
92:Q2	0.1587	0.1645	0.1525	0.1610	-0.0023	1.1719	0.9977	1.1746
92:Q3	0.1673	0.1688	0.1656	0.1600	0.0073	1.1821	1.0073	1.1735
92:Q4	0.1909	0.1880	0.1936	0.1657	0.0252	1.2104	1.0256	1.1802
93:Q1	0.2076	0.2016	0.2142	0.1685	0.0390	1.2307	1.0398	1.1835
93:Q2	0.2167	0.2162	0.2169	0.1723	0.0444	1.2419	1.0454	1.1880
93:Q3	0.2431	0.2450	0.2415	0.1825	0.0606	1.2751	1.0624	1.2002
93:Q4	0.2689	0.2649	0.2722	0.1881	0.0808	1.3085	1.0842	1.2069
94:Q1	0.3079	0.3114	0.3045	0.1890	0.1189	1.3605	1.1262	1.2080
94:Q2	0.3332	0.3349	0.3316	0.1936	0.1397	1.3955	1.1499	1.2136
94:Q3	0.3479	0.3527	0.3430	0.2027	0.1452	1.4161	1.1563	1.2247
94:Q4	0.3523	0.3602	0.3444	0.2108	0.1415	1.4223	1.1520	1.2347
95:Q1	0.3560	0.3548	0.3577	0.2287	0.1273	1.4276	1.1357	1.2570
95:Q2	0.3466	0.3395	0.3534	0.2454	0.1013	1.4143	1.1066	1.2781
95:Q3	0.3445	0.3437	0.3450	0.2514	0.0930	1.4112	1.0975	1.2859
95:Q4	0.3389	0.3397	0.3382	0.2575	0.0814	1.4034	1.0848	1.2937
96:Q1	0.3495	0.3573	0.3407	0.2643	0.0852	1.4184	1.0889	1.3026

LN_WRS = Logarithmic weighted repeat-sales index
LN_WRS_A = Logarithmic weighted repeat-sales index for random sample A
LN_WRS_B = Logarithmic weighted repeat-sales index for random sample B
LN_CPI = Logarithmic consumer price index
W(t) = Logarithmic real weighted repeat-sales index (LN_WRS – LN_CPI)
WRS = Non logarithmic weighted repeat-sales index
REAL_WRS= Non logarithmic real weighted repeat-sales index
CPI = Non logarithmic consumer price index

Appendix 3 Spurious Correlation in the WRS Index

Consider Case & Shiller's (1989) functional form (equation 5) where the log price P_{it} of the *i*th house at time t is given by $P_{it} = C_t + H_{it} + N_{it}$. Assume that there are only two observations on log housing prices. House *A* has an initial sale in period 0 and a subsequent sale in period 1, while house *B* has an initial sale in period 0 and a subsequent sale in period 2. The estimated changes in the log price index (using BMN or WRS procedure) since the number of observations equals the number of coefficients are:

For period 1:

$$P_{A1} - P_{A0} = C_1 - C_0 + H_{A1} - H_{A0} + N_{A1} - N_{A0}$$

For period 2:

$$-(P_{A1} - P_{A0}) + (P_{B2} - P_{B0}) = C_2 - C_1 - (H_{A1} - H_{A0} + N_{A1} - N_{A0}) + H_{B2} - H_{B0} + N_{B2} - N_{B0}$$

The index change between periods 0 and 1 is negatively correlated with the change between periods 1 and 2 because of common terms appearing with opposite signs.

There may also be positive serial correlation of estimated changes in the log price index. Assume there are three houses in the sample. House *A* sells in period 1 and 3. House *B* sells in period 0 and 2. House *C* sells in period 0 and 3. The estimated changes in the log index are:

For period 1, $(P_{C3} - P_{C0}) - (P_{A3} - P_{A1})$

For period 3, $(P_{C3} - P_{C0}) - (P_{B2} - P_{B0})$

These two changes are positively correlated in the model because house *C* has the same sign in both expressions whereas the specific changes to the other two houses are independent. This three-house case also demonstrates that there can be serial correlation between non-contiguous price changes.

This demonstrates that it is inappropriate to test efficiency of a housing sub-market by regressing real changes in the WRS index onto lagged changes and testing for significance of the coefficients. The same noise in individual house sales contaminates both dependent and independent and independent variables (Case & Shiller:1989:127-128).